

질의기반 사용자 프로파일을 이용하는 개인화 웹 검색

윤성희*

¹상명대학교 컴퓨터공학과

Personalized Web Search using Query based User Profile

Sung Hee Yoon*

¹Department of Computer Engineering, Sangmyung University

요약 사용자 입력 질의와 웹 문서에 포함된 단어들의 형태적 일치를 검사하여 관련 문서를 검색하는 검색엔진은 사용자의 개인별 관심 분야를 반영하는 검색 결과를 생성하기 어렵다. 본 논문에서는 개인별 관심사를 파악하여 질의 의도에 적합한 내용의 문서를 검색하는 개인화된 웹 검색 방법을 제안한다. 개인화 검색의 성능은 사용자의 개인적 관심사를 정확하게 표현하는 우수한 사용자 프로파일을 생성하는 전략에 좌우된다. 본 연구에서 개인 프로파일은 사용자가 최근 입력한 질의어들과 검색에서 클릭했던 문서들에 나타나는 주제어들이 출현 빈도를 반영한 가중치와 함께 등록된 데이터베이스이다. 특히 중의적 질의어의 정확한 의미를 결정하기 위해 워드넷을 기반으로 프로파일에 등록된 단어들과 의미 유사도를 계산한다. 기존 웹 검색 시스템의 사용자 측에 질의확장 모듈과 순위계산 모듈을 추가하는 확장모듈을 구축하여 비교 실험하였으며, 본 연구의 방법을 적용한 개인화 웹 검색의 결과는 특히 10위 이내 상위의 결과 문서들에 대해 92%의 정확률과 82%의 재현율을 보여 향상된 성능을 검증하였다.

Abstract Search engines that rely on morphological matching of user query and web document content do not support individual interests. This research proposes a personalized web search scheme that returns the results that reflect the users' query intent and personal preferences. The performance of the personalized search depends on using an effective user profiling strategy to accurately capture the users' personal interests. In this study, the user profiles are the databases of topic words and customized weights based on the recent user queries and the frequency of topic words in click history. To determine the precise meaning of ambiguous queries and topic words, this strategy uses WordNet to calculate the semantic relatedness to words in the user profile. The experiments were conducted by installing a query expansion and re-ranking modules on the general web search systems. The results showed that this method has 92% precision and 82% recall in the top 10 search results, proving the enhanced performance.

Keywords : personalized search, query expansion, re-ranking, user profile, WordNet

1. 서론

1.1 연구배경

정보검색 시스템은 사용자의 입력 질의(query)에 적합한 문서를 찾기 위해 대량의 문서들을 검색하고 질의 목적에 부합하는 문서들을 선별하고 순위를 매겨서 정답

후보 문서 집합으로 제시하는 소프트웨어 도구이다[1]. 정보검색 시스템의 성능은 사용자의 질의 의도에 적합한 문서들을 골라내고 신뢰할 만한 순위로 제공하는 기능에 기준을 두고 있다.

웹 검색에서 사용자들의 질의 행태를 분석한 연구들은 대부분 사용자들이 전문적이기 보다는 일반적이고 보편적

본 논문은 2014년도 상명대학교 교내연구비 지원의 결과임.

*Corresponding Author : Sung Hee Yoon(Sangmyung Univ.)

Tel: +82-41-550-5365 email: shyoon@smu.ac.kr

Received December 14, 2015

Accepted February 4, 2016

Revised (1st January 6, 2016, 2nd January 22, 2016)

Published February 29, 2016

인 용어들을 질의어로, 복잡한 검색 수식보다는 보통 두 개 정도의 간단한 단어로 질의한다고 보고하고 있다[2-3]. 반면에 사용자 검색 질의어의 의미를 정확히 판별하고 사용자들의 개인적인 관심 분야를 구분하는 기능은 검색 결과의 성능을 높이는데 매우 중요한 과정이다. 특히 질의어의 중의성(ambiguity) 문제는 단어의 형태적 일치를 기반으로 문서를 선별하는 방식에서 사용자의 차별적인 질의 의도가 검색 과정에 반영되지 못하는 중요한 이유가 된다. IT 기기나 소프트웨어 분야에 대해 관심이 많은 사용자의 검색어 'apple'이나 'virus'는 농산물이나 생물학에 관심이 많은 사용자의 검색어 'apple'이나 'virus'와는 의미가 분명히 다를 것이다. 이와 같이 검색 시스템의 성능 향상에 대한 사용자들의 기대는 점차로 개인의 관심사에 적합한 문서들만 결과로 제시하는 지능적인 기능의 품질 높은 맞춤형 검색으로 바뀌고 있다. 개인화된 웹 검색 기법은 정보검색의 개선된 기능으로서도 중요하지만 더 나아가 개인별 맞춤형 광고와 마케팅 등을 위해서도 매우 중요한 가치를 갖는 기술이 되고 있다.

1.2 연구내용 및 방법

본 논문에서는 검색 사용자들의 성향과 관심 분야를 검색 질의어와 참조했던 문서들의 주제어들로부터 추출하여 개인화된 프로파일(profile) 구성하고, 이를 활용하는 질의 확장(query expansion)과 순위 재계산(re-ranking) 과정을 통해 개인 관심사에 적합한 문서들을 검색하는 개인화 웹 검색(personalized web search) 방법에 대해 제안한다. 검색 시스템의 결과문서 집합이 사용자의 질의 의도에 얼마나 일치하는지에 대한 성능은 정확률(precision)과 재현율(recall)이라는 정량적 지표로 평가된다. 정확률은 결과로 제시된 문서집합 중에서 질의 의도에 적합한 문서의 비율이며, 재현율은 출현 기대하는 문서가 검색 결과에 포함되는 비율이라고 설명될 수 있다.

본 연구에서는 사용자들이 검색 시스템에 일정 기간 입력한 질의어들을 수집하여 사용자의 개인화된 선호도와 관심사를 반영하는 개인 프로파일 정보로 등록하고 질의 확장하여 검색 결과의 정확률을 향상시킨다. 또 검색된 후보 문서들을 관련 주제어들의 출현 중첩도를 이용하여 순위를 재계산하여 결과물을 생성한다. 특히 결과 문서들 중에서 사용자가 클릭하여 적극적으로 참조한 문서들에 대한 검색 행위 정보를 가중치로 반영하여 최신 정보를 유지한다. 이와 같은 과정을 질의확장 모듈과 순위재계산 모듈로

구현하여 실험하였으며, 검색 성능은 정확률과 재현율을 기준으로 실험을 통해 평가하였다.

2. 관련 연구

2.1 사용자 검색 경향 분석

정보검색 시스템은 기본적으로 입력 질의어와 웹 문서에 나타나는 색인어(index)의 형태적 일치를 검사함으로써 적합 문서 여부를 결정한다. 대부분의 사용자들은 아주 단순한 검색 질의어를 입력하고 검색된 문서들에 대한 관련성 평가를 통해 더 관련이 많은 문서를 찾기 위해 재검색하는 것이 일반적이다[3]. 검색 질의어의 길이는 평균 2.3개에 지나지 않으며, 복잡한 불린식(boolean expression) 등을 사용하는 경우가 드물고 대부분 단일 단어거나 복합명사의 형태로 입력한다고 한다. 결과적으로 사용자 질의의 정확한 의도를 파악하기 어려운 중의성이 자주 나타나고, 검색결과 집합에 대한 정확률과 재현율은 낮을 수밖에 없다. 또한 실험에 의하면 검색 시스템 사용자들은 일반적으로 상위에 제시되는 몇 개 문서들만을 주로 참조하는 경향이 있으므로 검색 시스템의 정확률이 재현율에 비해서 더 중요하다고 평가된다.

2.2 개인 관심사 및 선호도

사용자의 개별적인 선호도나 관심사에 대한 정보의 집합인 프로파일은 특정 사용자에게 연관된 개인화 데이터의 모이며, 개인의 차별성에 대한 명시적인 디지털 표현이다. 사용자가 관심분야를 직접 입력하게 하는 정적인 수집 방식이나 사용자가 이전에 검색했던 페이지에 대한 정보를 지속적으로 수집하여 검색에 반영하는 동적인 수집 방법이 사용된다. 개인 관심사를 표현하는 주제어들을 수집하고 그들 간 개념 네트워크(Concept Network)로 생성하여 개인 프로파일로 사용하는 연구들이 소개되고 있다[4-7].

공통의 관심 분야를 갖는 사용자 집단의 성향을 프로파일 정보에 반영하는 연구들에서는 포털의 개인 블로그나 가입 활동하는 카페 등 커뮤니티(communitiy)를 기반으로 정보를 수집한다[8-9]. 개인의 관심사를 다른 사용자들의 관심사와 결합하여 다수 사용자의 성향에 대한 정보를 표현할 수 있는 연결망 구조로 확장하는 방법을 소개했다. '사람들에 의한 분류법'라 불리는 폭소노미

(folksonomy)의 태그를 개인화 검색에 활용하는 연구에서는 태그들을 중심으로 문서의 클러스터링 과정을 거치며, 문서의 태그가 사용자의 질의어로 사용되는 경우가 많다는 특성을 활용하는 방법이다[10].

사용자가 입력한 짧은 질의어는 정확한 검색 결과를 얻기에 정보가 부족하므로, 검색 시스템이 사용자의 질의어와 관련이 깊다고 판단되는 연관 검색어들을 추가로 추천하고 사용자가 이를 선택하게 하거나, 사용자의 검색 반응 정보(relevant feedback)를 활용하는 방법도 제안되었다[11-12]. 이와 같은 방법들은 검색 과정에서 사용자가 수동으로 연관된 질의어를 선택하는 단계를 거쳐야 하고, 태그를 활용하거나 연관 질의어를 추천하기 위해 외부 집단의 정보를 참조하는 시스템으로 구현되어야 한다. 텍스트 뿐만 아니라 이미지나 영상 등 다양한 형식의 데이터의 분석을 통해 가치 있는 정보를 얻어내는 빅데이터에서도 유사도와 빈도수를 계산하는 알고리즘을 적용한다. 텍스트 데이터 처리에는 사전정보를 이용하고 불용어 처리와 형태소 분석 등 다양한 자연어처리 결과를 활용하며 질의어나 주제어들 간의 형태적 유사도와 빈도를 중심으로 처리하거나 온톨로지 등의 대용량 자료를 활용하기도 한다[13-14]. 본 연구는 워드넷을 이용하여 질의어와 주제어들의 사이의 의미 유사도(semantic similarity)를 계산하고 순위를 재계산하는 개인화 검색 시스템을 소개한다.

2.3 질의어와 주제어의 의미체계

사용자의 질의 기반의 웹 검색에서 해결해야 하는 중요한 문제 중 하나는 중의적 단어에 대해 사용자의 의도에 맞는 정확한 의미를 결정하는 과정이다[15]. Fig. 1은 애플사의 제품이나 관련 기기에 관심이 많은 사용자와 여행이나 음식에 관심이 많은 사용자가 주로 질의하고 검색하는 문서들의 주제어 그룹을 비교한 것이다. 중의성 해소를 위한 연구는 온톨로지(ontology)나 시소러스(thesaurus) 등 다양한 방법을 적용한다. 대용량 말뭉치에서 추출한 통계적인 정보에 기반한 방법은 기계학습(machine learning)을 위해 개별 의미를 부착한 대용량의 말뭉치를 생성하고 감독(supervised) 또는 비감독(non-supervised) 중의성 해소 기법을 적용한다. 하지만 언어자원으로서 신뢰할 만한 대규모 의미부착 말뭉치는 생성하는 비용과 시간이 절대적으로 많이 드는 단점이 있다.



Fig. 1. Personal topic words cloud

대용량 말뭉치 대신 단어 중의성을 해결하기 위해 워드넷(WordNet)이라는 지식베이스가 여러 연구에서 활용되고 있다. 워드넷은 미국 프린스턴(Princeton) 대학에서 개발되고 공개되어 영어권의 어휘 의미 체계 연구에 널리 활용되고 있으며, 계층적 방법으로 어휘들의 의미 연관관계를 체계화하였다[16-17]. 자연어처리 연구 분야에서 자주 이용되는 대형 어휘 정보 데이터베이스이며 명사, 동사, 형용사와 부사를 각각 동의어 집합으로 분류하고 별개의 개념인 신셋(synset: synonym set)이라는 단위를 노드로 표현한다. 단어들 사이의 개념과 연관 관계를 잘 표현하고 있으므로 기계번역에서 대역어나 대응어 선택 등을 위해 유용하게 활용된다. Fig. 2는 워드넷에서 의미 계층의 한 가지 예이다.

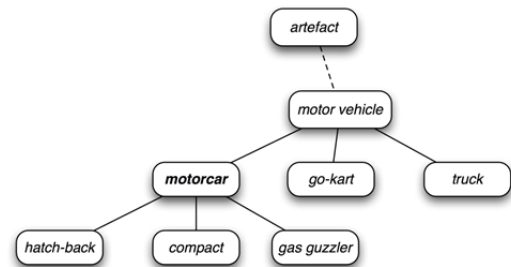


Fig. 2. Example of WordNet semantic network

영어 워드넷(PWN: Princeton WordNet)을 기반으로 세계 여러 언어에 대한 워드넷을 확장하는 연구가 활발히 진행되고 있으며, 주로 영어 워드넷을 대역하여 구축한 후 정제하는 방법으로 진행되고 있다[18-21]. 본 연구에서는 워드넷을 이용하여 질의어와 관심 주제어들 사이의 의미 근접 정도를 계산하여 사용자의 관심분야를 결정하는데 활용한다.

3. 질의기반 개인화 검색

본 연구에서 제안하는 개인화 검색모델의 처리 과정은 Fig. 3과 같이 요약될 수 있다. 사용자의 입력 질의를 확장하는 선처리 과정과 검색 결과에 대한 순위를 재계산하는 후처리 과정이 개인화 검색 과정의 핵심이다. 연관 검색어 추천 시스템과 비교하면 사용자 질의 입력과 문서 클릭 등의 암시적인 검색 행위를 기반으로 주제 관련 검색어를 자동 선별하여 질의를 확장하는 방식이며, 중의적 질의어와 문서 주제의 의미를 판별하기 위해 워드넷을 기반으로 의미 유사도를 계산하고 사용자가 참조했던 문서들의 주제어들을 수집하여 개인 프로파일을 갱신한다.

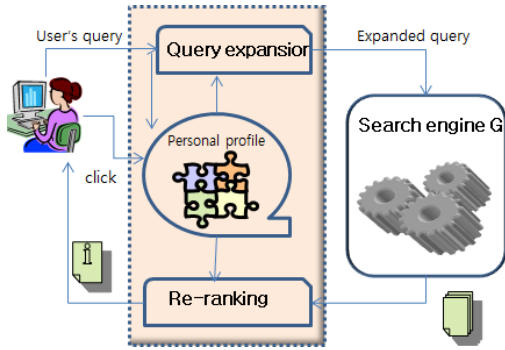


Fig. 3. Data flow of personal web search

3.1 질의어 중의성 해소

중의성을 갖는 주제어들의 정확한 의미는 사용자의 관심을 반영하는 다른 주제어들과의 의미적 근접성에 의해 결정될 수 있다. 워드넷에서 노드 간의 거리는 어휘의 의미 거리이며 개념 간 유사도를 판단하는 정량적인 값이다. 워드넷에서 의미 유사도를 계산하는 알고리즘은 여러 가지가 소개되고 있으나 간단하면서도 가장 많이 적용되는 방법은 경로 길이(path length)에 기반한 계산 방법이며 본 연구에서도 이와 같은 방법을 적용하였다 [22-24]. 사용자들 개인별 관심 분야에 따라 입력하는 질의어의 집합이 다르며, 중의적 질의어에 대한 정확한 질의 의도는 이 집합에 속해있는 단어들의 그룹과 의미 근접성에 기반을 두고 결정하는 방법이다. 워드넷의 두 노드 s 와 t 사이 거리의 가까운 정도 $Sim(s,t)$ 는 다음의 식 (1)과 같다. 여기서 $Distance(s,t)$ 는 노드 s 와 t 사이 경로의 거리를 말한다.

$$Sim(s,t) = 1 / Distance(s,t) \quad (1)$$

3.2 개인화 프로파일 구축

본 연구에서의 개인화 정보, 즉 사용자 프로파일은 사용자가 입력한 질의어들과 실제 클릭했던 문서들로부터 수집되는 주제어들의 집합이다. 검색 행위가 반복되면서 최근의 관심 분야 질의어를 중심으로 가중치가 갱신되며, 질의어가 중의성을 갖는 경우에는 워드넷을 기반으로 주제어들의 의미 유사도를 계산한다. 사용자의 프로파일을 구성하는 정보들은 다음과 같으며, Fig.4는 개인화 검색에서 사용자 프로파일의 위치와 역할을 보여준다.

사용자 개인화 프로파일 =

$$\begin{aligned} & \{ \text{사용자의 최근 입력 질의어} \} \\ & \cup \{ \text{사용자가 클릭한 문서의 주제어} \} \\ & \cup \{ \text{워드넷에 기반한 주제어들의 의미} \} \\ & \cup \{ \text{빈도에 기반한 주제어들의 가중치} \} \end{aligned}$$

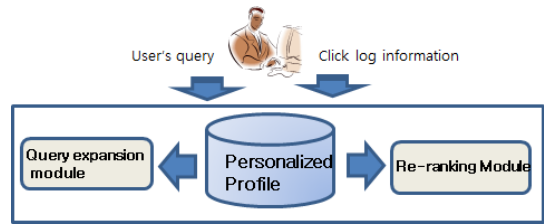


Fig. 4. Personalized Search with query expansion and re-ranking

3.3 질의확장과 순위 재계산

사용자 프로파일은 개인화 검색 과정에서 두 가지 방법으로 반영된다. 사용자의 단순한 입력 질의어를 프로파일에 등록된 최근 관심 주제어와 함께 확장하는 과정과 검색된 결과 문서들을 관심도가 높다고 판단되는 순위로 재계산하여 보여주는 과정이다. 사용자의 질의어는 프로파일에 등록된 높은 가중치의 주제어와 함께 두 개의 질의어 또는 세 개의 질의어로 확장되어 일반 검색엔진에 입력된다. 이와 같이 검색된 결과 문서 집합들 중에서 사용자가 실제 클릭하여 참조하는 행위는 해당 문서에 대한 적합성을 보여주는 묵시적 평가이다. 특히 사용자들이 상위의 10개 내외의 문서들만을 중심으로 참조한다는 연구 결과에 따르면 검색 문서들의 순위를 결정하는

것은 검색 성능의 평가에 매우 중요한 과정이다. 본 연구에서는 순위 재계산을 위해 사용자 프로파일의 관심 주제어들과 검색결과 문서에 출현하는 주제어들과의 중첩 정도를 계산한다. 또한 실제로 사용자가 클릭하는 문서에 대해서는 관심분야 주제어 갱신과 가중치 갱신이 이루어진다. 확장 질의어로 검색된 문서집합 D_{qe} 에 포함된 문서들의 순위 값은 식 (2)와 같이 계산된다. $d_q r$ 은 D_{qe} 중 순위가 r 인 문서이다.

$$T(d_q^r) = 1/\log(r+1) \quad (2)$$

개인 프로파일에 나타나는 상위 주제어들로 질의 확장 후 검색 결과에 대한 중첩도를 순위 재계산에 반영하는 방법은 아래 식 (3)과 같다. 여기서 d_{qi} 는 i 번째 주제어로 확장된 질의어 q_i 로 검색된 문서집합을 의미한다.

$$T(d_q) = \sum_{i=1}^k T(d_{qi}^r) \quad (3)$$

4. 실험 및 평가

본 연구에서 제안한 개인화 검색 방법은 기존 검색 시스템을 사용하는 사용자 측에 확장 모듈(expanded module) 방식으로 구현되었으며, Fig. 5는 검색 시스템의 내부 구성도이다.

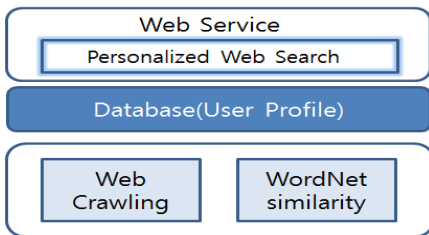


Fig. 5. Internal hierarchy of personal web search

음악, 여행, 컴퓨터, 프로그래밍, 스포츠 등 다양한 성향과 관심분야를 가진 대학생 사용자 20명을 대상으로 실험하였으며, 사용자들의 초기 프로파일은 일주일 동안의 입력 질의어들과 실제 참조 문서들의 주제어들을 기반으로 구성하였다. 실험 환경과 결과에 대한 일반적인 요약은 Table 1과 같다.

Table 1. Test summary of personalized search

number of users	20
average number of tested queries	23.4
maximum length of query expansion	3
size of tested and ranked documents	50
maximum rank movement	7

성능 평가 방법은 검색 성능에 대한 정량적 척도인 정확률과 재현율을 구하기 위해 개인화 검색의 결과가 각 사용자가 실제로 관심을 갖는 문서들인지를 판별해야 하므로 사용자들이 직접 수동으로 평가를 하였다. 정확률과 재현율은 다음 (4)와 같이 계산되며, 여기서 R 은 관련 문서들(Relevant documents)의 집합이고, C 는 검색 문서들(Collected documents)의 집합이다. Table 1에서 보듯이 본 실험에서 $|C|$ 는 50이다.

$$\text{정확률(precision)} = |R \cap C| / |C| \quad (4)$$

$$\text{재현율(recall)} = |R \cap C| / |R|$$

개인화 검색의 성능 향상 정도를 평가하기 위해 확장 모듈을 적용하지 않은 기존의 검색 시스템의 검색 결과에 대한 평가를 비교하였다. 실험의 결과는 Table 2와 Fig. 6 및 Fig. 7과 같이 요약된다. 정확률과 재현율로 측정된 개인화 검색 시스템의 검색 결과에 대한 만족도는 상대적으로 크게 향상되었으며, 특히 상위 10~20개 문서에 대한 사용자의 만족도가 높게 나타났다. 대부분 사용자들이 상위의 문서들만을 중심으로 관심을 보이고 클릭한다는 검색 행태에 대한 연구 결과를 반영한다면 이 결과는 매우 의미가 있다고 할 수 있다. 반면에 20~30위나 그 이하 순위의 문서들에 대한 성능 차이의 의미는 상대적으로 적다고 본다.

Table 2. Precision and recall evaluation

rank	precision		recall	
	general search	personalized search	general search	personalized search
1~10	0.75	0.92	0.62	0.82
11~20	0.50	0.84	0.25	0.47
21~30	0.47	0.57	0.20	0.25
31~40	0.38	0.41	0.20	0.21
41~50	0.35	0.42	0.15	0.15

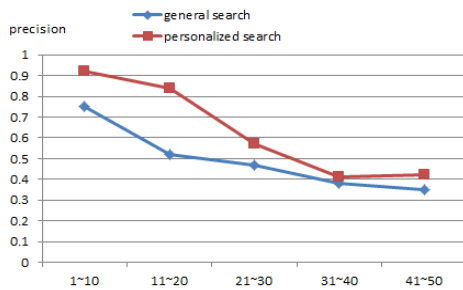


Fig. 6. Evaluation graph of precision

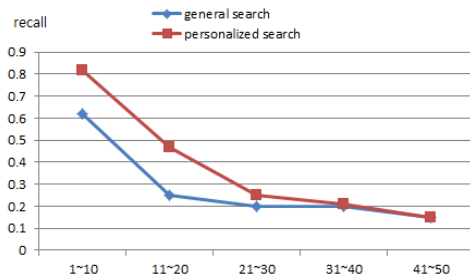


Fig. 7. Evaluation graph of recall

5. 결론

본 연구는 개인의 관심사를 파악하여 차별화된 검색 결과를 제공하는 개인화 웹 검색 방법을 제안하였다. 개인의 관심 주제를 수집하여 프로파일 데이터베이스에 저장하고 지속적으로 갱신하여 최신 관심사를 유지하면서 검색 결과를 제공하는 것이 목적이며, 사용자의 검색 환경에서 확장모듈로 구축되는 방법이다. 사용자 프로파일을 구성하는 내용은 사용자가 최근에 입력한 질의어들과 최근 실제 검색했던 문서들의 주제어들이며, 특히 중의적 성격을 갖는 주제어들의 의미를 판별하기 위해 워드넷을 기반으로 사용자의 관심을 반영하는 주제어들 사이의 의미 유사도를 계산하였다. 개인 관심 분야가 다른 실제 사용자들을 대상으로 개인화 검색과 기존 검색을 실험하고 결과를 비교 평가하였으며, 검색 결과 상위 10위 이내의 문서들을 중심으로 92%의 정확률과 82%의 재현율로 성능이 향상된 결과를 보였다.

개인의 관심 분야와 선호도를 고려하는 맞춤형 검색 시스템의 개발은 광고나 마케팅 등 다양한 시장으로 넓은 활용이 이루어질 것으로 기대된다. 사용자의 만족도

를 높이는 우수한 품질의 검색 결과를 생산하기 위해 질의어와 웹 문서 주제어들의 중의성 해소나 문서 분류 등 자연어 처리 분야의 여러 연구 결과들을 적극적으로 반영하는 지속적인 연구는 정보검색에 큰 발전을 가져올 것으로 기대된다.

References

- [1] Baeza-Yates Ricardo, Reberio-Neto Berthier, Modern Information Retrieval. Addison Wesley, 1999.
- [2] Soyeon Park, Joon -Ho Lee, "Investigating Web Search Behavior via Query Log Analysis," Journal of the Korean Society of Information Management, Vol. 24, No. 4, pp. 255-265. 2002.
- [3] Soyeon Park, Joon-Ho Lee, Ji Seoung Kim, "An Analysis of Query Types and Topics Submitted to Naver," The Journal of the Korean Society for Library and Information Science, Vol. 39, No. 11, pp. 265-278, 2005.
DOI: <http://dx.doi.org/10.4275/KSLIS.2005.39.1.265>
- [4] S. Koratkar, S. A. Takale, "Deriving Concept Based User Profile for Search Engine Personalization," International Journal of Science and Research, Vol.4, No. 6, pp.3086-3089, 2013.
- [5] K. W. Leung, W. Ng, D. L. Lee, "Personalized Concept-Based Clustering of Search Engine Queries,," IEEE Transactions on Knowledge and Data Engineering, Vol.20, No. 11, pp.1505-1518, 2008.
DOI: <http://dx.doi.org/10.1109/TKDE.2008.84>
- [6] N. Sharma, M. Sharma, O. J. Gupta, "Search Engine Personalization Using Concept Based User Profiles," International Journal of Scientific Research Engineering & Technology(IJSRET), Vol. 2, No. 4, pp.084-087, 2012.
- [7] K. R. Remesh Babu, P. Samuel, "Concept Networks for Personalized Web Search Using Genetic Algorithm," International Conference on Information and Communication Technologies (ICICT 2014), pp. 566-573, 2015.
- [8] Taebok Yoon, Seungjoon Lee, Kwangho Yoon, Jeehyong Lee, "Design and Application of Multi Concept Model based on Web-using Information," Journal of Internet Computing and Services, Vol. 10, No. 5, pp.95-104, 2009.
- [9] Gunwoo Park, Sang-Hoon Lee, "Personalized Search based on Community through Automatic Analysis of Query Patterns," Journal of the Korean Institute of Information Scientists and Engineers: Database, Vol. 36, No. 4, pp. 321-326, 2009.
- [10] Dongwook Kim, Sooyong Kan, Hanjoon Kim, Byungjeong Lee, "Folksonomy-based Personalized Web Search System," Journal of Digital Contents Society, Vol. 11, No. 1, pp. 105-116, 2010.
- [11] Soyeon Park, "Analysis and Evaluation of Term Suggestion Services of Korean Search Portals: The Case

of Naver and Google Korea," Journal of the Korean Society of Information Management, No. 20, Vol. 2, pp.297-315, 2013.

DOI: <http://dx.doi.org/10.3743/kosim.2013.30.2.297>

- [12] Sung Hee Yoon, "Using Query Word Senses and User Feedback to Improve Precision of Search Engine," Journal of the Korean Society of Information Management, Vol. 26, No. 4, pp. 81-91, 2009.
DOI: <http://dx.doi.org/10.3743/kosim.2009.26.4.081>
- [13] Jeong Heo, Pum-Mo Ryu, Yoon Jae Choi, Hyun Ki Kim, and Cheol Young Ock, "An Issue Event Search System based on Big Data for Decision Supporting: SocialWisdom," The Proceedings of Korean Institute of Information Scientists and Engineers Vol. 39, No. 5, pp.381-394, 2013.
- [14] Boon Hee Kim, "Words Recommendation Algorithm for Similarity Connection based on Data Transmutability," The Journal of The Korea Institute of Electronic Communication Sciences, Vol. 8, No. 11, pp. 1719-1724, 2013.
DOI: <http://dx.doi.org/10.13067/JKIECS.2013.8.11.1719>
- [15] Yong-Gu Lee, Young-Mee Chung, "An Experimental Study on an Effective Word Sense Disambiguation Model Based on Automatic Sense Tagging Using Dictionary Information," Journal of the Korean Society for Information Management, Vol.24, No. 1, pp. 321-342, 2007.
DOI: <http://dx.doi.org/10.3743/kosim.2007.24.1.321>
- [16] Wikipedia,
<http://en.wikipedia.org/wiki/WordNet>
- [17] WordNet: A Lexical Database for English.
<http://wordnet.princeton.edu/>
- [18] KIBS : Korean Information Base System.
<http://kibs.kaist.ac.kr>
- [19] Young-Bum Kim, Yu-seop Kim, "A Question Example Generation System for Multiple Choice Tests by utilizing Concept Similarity in Korean WordNet," Journal of the Korean Information Processing Society, Vol. 15(A), 2008.
- [20] Kyeong-Kook Park, Kwang-Mo Lee, Yu-Seop Kim, "WordNet Extension for IT Terminology Using Web Search," The Proceedings of Korean Institute of Information Scientists and Engineers Conference, pp. 189-193, 2007.
- [21] In Keun Lee, Dosam Hwang, Younggyun Hahm, Key-Sun CHoi, "Open Korean WordNet(KWN): Dictionary-based Semi-Automatic Development," The 26th Annual Conference on Human & Cognitive Language Technology, 2014.
- [22] Mi-Young Cho, Jun-Ho Choi, Pan-Koo Kim, "Design of Conceptual Image Annotation System Using WordNet," The Proceedings of the Korea Multimedia Society Conference. pp. 1081-1086., 2002.
- [23] L. Meng, R. Huang, J. Gu, "A Review of Semantic Similarity Measures in WordNet," International Journal of Hybrid Information Technology. Vol. 6, No. 1., 2013.
- [24] T. Simpson, T. Dao, "WordNet-based semantic similarity measurement," 2010,
<http://www.codeproject.com/Articles/11835>.

윤 성 희(Sung Hee Yoon)

[정회원]



- 1989년 2월 : 서울대학교 대학원 컴퓨터공학과(공학석사)
- 1993년 8월 : 서울대학교 대학원 컴퓨터공학과(공학박사)
- 1994년 3월 ~ 현재 : 상명대학교 컴퓨터공학과 교수
- 1999년 1월 ~ 2000년 1월 : University of Michigan 방문연구 교수
- 2007년 8월 ~ 2008년 7월 : University of Victoria 방문 연구교수

<관심분야>

자연어처리, 한국어정보처리, 정보검색